Análisis de voz, aplicaciones en sistemas de entrenamiento y rehabilitación.

I Dr. Sergio Suárez Guerra.
Investigador Titular ICIMAF, Cuba.
Profesor Investigador CIC – IPN, México.
e-mail: ssuarez@cic.ipn.mx **Telf: 5729-6000 ext. 56588.**
e-mail: sergio@cidet.icmf.inf.cu **Telf; 32-4085.**
II Lic. Liseth García Robles.
Estudiante Maestría CIC – IPN, México.
VIII Congreso Internacional de Informática en la Educación

Resumen.

Las aplicaciones computacionales al procesamiento de voz, fueron dirigidas en primera instancia a la solución del problema de comunicación directa entre la computadora y el usuario, desarrollándose sistemas para el reconocimiento de voz, interprete de comandos, conversión texto - voz y por último en fase de desarrollo la traducción automática.

De forma sostenida la computación ha sido introducida en diversos aspectos de la medicina, mediante la creación de equipos de diagnóstico computarizados que disminuyen considerablemente los errores humanos en análisis de laboratorio, creación de medicamentos y análisis de señales y sistemas fisiológicos.

Se presentan los resultados de un proyecto en ejecución, que ha ganado muchos adeptos en el análisis de las características, métodos y formas de enfocar el análisis y presentación del procesamiento de la voz, así como los resultados alcanzados en este caso.

Palabras Claves: Procesamiento de voz, Parámetros de voz, voz y habla.

Antecedentes.

Las patologías de Voz, Habla y Auditivas, se consideran los principales problemas en la comunicación humana. Una cifra proporcionada por especialista en esta disciplina, sitúa el problema con un índice de incidencia global del 10% de la población. Las clasificaciones son diversas atendiendo al grado de severidad de la dolencia.

1 Introducción.

El procesamiento de la voz, con el propósito de realizar la comunicación entre dispositivos automáticos y el ser humano utilizando el lenguaje oral, ha sido motivo de gran expectativa y el sueño de muchos investigadores durante las últimas tres décadas (Rabiner, 1989; Rabiner y Juang, 1993; Reyes, 1994; Savage, 1995; Barrón, 1998; Barrón y otros, 1999; CVRecVoz, 1999). Lo que parecía ser un problema sencillo, con el tiempo se ha ido convirtiendo en una tarea cada vez más complicada. La audición no está delimitada solamente al órgano sensorial periférico u oído, sino que es una función mucho más compleja de elaboración neurológica. El oído cumple la función de **"micrófono",** capaz de captar el sonido,

pero la elaboración y reconocimiento del mismo es una función puramente cerebral. El oído humano es capaz de percibir las vibraciones sonoras en frecuencias comprendidas entre los 16 y 18.000 Hz. Estas vibraciones son captadas por el pabellón de la oreja y transmitidas por el conducto auditivo externo al tímpano, el cual al vibrar pone en movimiento a la cadena de huesecillos del oído medio. Estos transmiten la vibración hacia los líquidos del oído interno lo que provoca la estimulación del órgano de Corti (porción auditiva del oído interno), allí se genera una actividad eléctrica que toma las características del sonido que lo ha producido (Neurofisiología, 1998; Tinnitus). La cuestión más importante y difícil de resolver, es la naturaleza multi e interdisciplinaria del problema: captación de señales, preprocesamiento, estadística matemática, redes neuronales, lógica difusa, reconocimiento de patrones, lingüística, fisiología, ciencias de la computación, etc. Esto obliga a concebir los trabajos de procesamiento de voz desde varios puntos de vista, teniendo en cuenta la aplicación de los resultados: compresión de voz (reducir el número de bits para almacenar o transmitir voz, incrementando la razón de calidad - compresión para robustecer la codificación en canales no ideales de transmisión); reconocimientos de voz (dependiente e independiente del locutor); reconocimiento del locutor (identificación o certificación del locutor) (Reddy, 1966; Furui, 1997), dentro de esto tenemos algo muy de moda, como es la traducción automática de idiomas (Wahlster, 1996); síntesis de texto – voz y por último análisis de voz (forma de onda, parámetros característicos).

La tarea de análisis de voz constituye la base para el desarrollo y entendimiento de la producción y síntesis de voz, así como de los algoritmos para la identificación, clasificación y rehabilitación de patologías en la producción de la Voz, Habla y Lenguaje articulado.

Dentro de la temática de análisis de voz se han desarrollado los sistemas para el entrenamiento y aprendizaje del habla, los cuales se basan en la representación de los perfiles acústicos espectrales correspondientes a frases o palabras (Charles, 1986; EXPARAM, 1996; Oppenheim,1996; Kay, 1998; Voice) o la animación de imágenes (Speech Viewer,1996) , según sea la modalidad elegida. El idioma tiene una fuerte influencia en la selección de parámetros básicos, en cuanto a reconocimiento de voz se refiere, pero para problemas de articulación pueden no ser tan dependientes.

Las herramientas utilizadas para el reconocimiento de voz e identificación de locutor son similares, lo que diferencia su alcance es la aplicación (Mansfred,1981; Havawaka,1997; Furui, 1997; Suárez, 2000).

2 Métodos de análisis acústico.

2.1 Análisis oscilográfico

El oscilograma: nos permite hacer la representación de las variaciones de amplitud en el habla a lo largo del tiempo como el análisis de la sonoridad, la duración, las pausas, el acento y el ritmo.

2.2 Análisis espectral

Para ello se utiliza la FFT, *Fast Fourier Transform* - Transformada Rápida de Fourier con lo que obtenemos un

Espectro: representación de la frecuencia y la amplitud de los armónicos en un instante de la señal sonora del habla.

Análisis de la sonoridad y de la estructura formántica (timbre).

Oscilograma del enunciado " esto es una demostración". Espectro de la vocal [e]

2.3 LPC, Linear Predictive Coding - Codificación por predicción lineal

Es representación de la frecuencia y la amplitud de los picos espectrales que representan resonancias del tracto vocal en un instante de la señal sonora del habla. Análisis de la estructura formántica (timbre), *(de http://www.sci.fi/~pitchsys/index.html).*

2.4 Análisis espectrográfico

Espectrograma: representación de las variaciones de la frecuencia y la intensidad en el habla a lo largo del tiempo nos ayuda en el análisis de la sonoridad, la duración, la estructura formántica (timbre), la intensidad, las pausas, el acento y el ritmo.

Espectrograma del enunciado "esto es una demostración". Espectro LPC de una vocal.

2.5 Análisis melódico

Curva melódica: representación de la variación de la frecuencia fundamental del habla a lo largo del tiempo. Análisis de la melodía, el acento y la entonación.

2.6 Análisis de intensidad

Curva de intensidad: representación de la variación de la intensidad del habla a lo largo del tiempo

Análisis de la intensidad, el acento, el ritmo y las pausas.

Curva de intensidad del enunciado "signalyze"
(de http://agoralang.com/signalyze.html).

Curva melódica del enunciado "te espero en la calle"
(de Estruch, Garrido, Llisterri y Riera, 1996)

3. Línea de investigación en desarrollo.

Tomando como antecedente los trabajos realizados en EXPARAM V.1.2, se trabaja en un proyecto de investigación "Sistema de procesamiento de voz, para aplicaciones médico foniátricas", que abarca diferentes objetivos.

En la siguiente figura se pueden apreciar claramente los módulos que conforman el proyecto, los cuales a su vez, están enmarcados en diferentes secciones, constituyendo cada una de ellas, una aplicación u objetivo.

La tesis 1, es un sistema para la "Gestión y análisis acústico en consultas de foniatría", el cual servirá para darle seguimiento y tratamiento de rehabilitación a pacientes con trastornos del habla. El proceso de extracción y representación de perfiles acústicos a partir de parámetros de la voz, se considera un análisis cualitativo, ya que lo que se busca es retroalimentar al paciente y al médico especialista con gráficas que reflejen la variación de los parámetros seleccionados a lo largo del tiempo de producción de la voz, así como su comparación morfológica y en ningún momento se realizan cálculos numéricos en la desviación de las representaciones gráficas.

La tesis 2 es una extensión de la tesis 1, donde se pretende realizar el análisis cuantitativo de la extracción de los parámetros acústicos de la señal voz, lo cual servirá para realizar la clasificación diagnóstica de algunos pacientes a partir de estas mediciones.

La tesis 3, tiene como objetivo realizar una medición de la funcionalidad del oído, mediante técnicas de audiometría. Los resultados de estas mediciones se depositan en los datos de la historia clínica de los pacientes que contiene el sistema.

Una aplicación independiente es la de aplicaciones para la Educación Especial, el cual se alimenta de varios de los módulos del sistema. Con esta aplicación se pretende dotar a las escuelas especiales y a los alumnos y pacientes, de un sistema autónomo para el entrenamiento y la rehabilitación de problemas de voz, sin necesidad de estar asistiendo a consultas o clases.

Hay un módulo que no está enmarcado dentro de las tesis, pero del cual hacen uso todos y es el módulo de representación de imágenes GIF, con el cual y como una ayuda adicional, se presenta una realimentación visual del objeto que representa la palabra en proceso de análisis.

4. Gestión y análisis acústicos en consultas de foniatría.

El objetivo principal del trabajo realizado es disponer de un sistema para le gestión de pacientes en consulta de foniatría y realizar el análisis de voz por métodos computarizados. Con respecto a esto último es posible realizar la captura y almacenamiento de las señales de voz de los pacientes y representar los perfiles de los parámetros de palabras y frases de las mismas, así como comparar los perfiles obtenidos en la consulta con los resultados del procesamiento de un corpus de voces de personas normales, al cual el sistema tiene acceso.

Sistemas que realicen tanto la gestión de consulta como el análisis acústico de la señal voz y realizados en el país no existen, de ahí la importancia de este trabajo.

Las ventanas principales de este trabajo se muestran a continuación.

Módulos componentes de la aplicación.

Formulario para el llenado de los datos del paciente.

En la primera ventana se escoge el tipo de tarea a realizar con el paciente y la segunda es para el caso de llenado del expediente médico.

Una vez llenado el expediente médico existe la posibilidad de obtener y guardar en archivos la voces del paciente, así como realizar el análisis acústico de la señal de voz recibida. En la siguiente figura podemos ver una de estas representaciones.

Parámetro de energía de la palabra “lámpara”.

Las condiciones de tiempo de presentación y la selección del perfil paramétrico a observar lo decide el usuario del sistema con el menú de la barra superior.

Es posible oír el sonido que está representado en pantalla, lo cual proporciona una doble realimentación al especialista y paciente.

5. Aplicación educativa.

Un sistema con nuevas características está en vías de aplicación. Donde la presentación de perfiles acústicos se realiza a partir de ambientes en programación visual y en tiempo real; además de combinar las características de ser un sistema orientado al registro de pacientes, ofrecer menúes de rehabilitación y entrenamiento.

En esta pantalla se observa la visualización de una señal de voz de la palabra “espada”, la misma puede ser reproducida de forma auditiva para ser escuchada o ser sometida al procesamiento digital, para que los perfiles acústicos de los diferentes parámetros sean presentados, analizados y comparados por el usuario del sistema.

En la siguiente pantalla es posible observar la extracción de dos perfiles acústicos de la señal voz: en color rojo vemos el parámetro de intensidad (A0) y en color amarillo el de las frecuencias mayores de 4000 Hz (RO).

Con esta representación es posible observar la ocurrencia del fonema /s/ en la parte inferior y determinar si el fonema ha sido adecuadamente producido de forma inequívoca.

En pantalla inferior siguiente vemos la representación paramétrica de la palabra “campana”, donde es posible observar la ausencia del parámetro F2 (segundo formante), para los fonemas /m/ y /n/.

También es posible presentar en cada una de las ventanas el resultado de la extracción del mismo parámetro para dos sonidos diferente, con el fin de comparar visualmente la diferencia del mismo para los dos sonidos. Esta presentación es muy utilizada para la rehabilitación y el entrenamiento de pacientes.

Los perfiles de parámetros acústicos que se diseñaron para su extracción y representación con este sistema incluyen, además de los mostrados: formante F1; cruce de ceros; tono fundamental y sus variaciones; espectro de frecuencia; y acotación por cursores de la zona de análisis y representación

Los sonidos de las representaciones de las ventanas superiores e inferiores, pueden ser escuchados con la ayuda del menú de Reproducción y botones que se insertan en la pantalla principal.

3 Conclusiones.

El trabajo realizado hasta el momento para el procesamiento de voz en aplicaciones médicas y educativas, ha reflejado la indiscutible necesidad de vincular el conocimiento de la Fonética Acústica y su vinculación con la fisiología de los sistemas productores de la voz y auditivos.

Los métodos utilizados para el estudio y análisis de la producción de voz, así como de la percepción e interpretación de la misma, son el resultado de la evolución del sistema nervioso central y las relaciones sociales entre los hombres.

Es imposible hoy en día no tener en cuenta el empleo de la técnica de computación, para resolver los problemas de diagnóstico rehabilitación, reconocimiento de voz, traducción automática y todo aquello que está ligado a ambos sistemas.

Se requiere del desarrollo de sistemas de cómputo que incorporen la experiencia médica y el análisis de señales en perfecta armonía, para que de una manera multidisciplinaria y eficiente, se vayan resolviendo las hipótesis, problemas y necesidades que cada día surgen alrededor del tema.

Referencias

Barrón Ricardo. Reconocimiento de palabras aisladas usando cuantización vectorial. Tesis de Maestro en Ingeniería de Cómputo. CIC – IPN, D.F. México, octubre de 1998.

Barrón Ricardo, Sergio Suárez, Claudia Montezuma. Reconocimiento de comandos verbales utilizando cuantización vectorial y redes neuronales. CIC - IPN. México. Informe técnico: Serie roja No. 40. ISBN 970-18-2673-6. 1999.

Charles Sterling Williams. Designing Digital Filters. Englewood Cliff, New Jersey. Prentice Hall, c 1986.

CVRecVoz; Sistema para el Reconocimiento de Palabras Aisladas utilizando Cuantización Vectorial y Técnica Multisección. Manual de Usuario. Laboratorio de Tiempo Real CIC – IPN. D.F. México. Junio 1999.

EXPARAM V.1.2. **Suárez Sergio.** Sistema para la extracción y análisis de parámetros de la voz. Manual de Usuario. Instituto de Cibernética, Matemática y Física (ICIMAF), Cuba. 1996. Está disponible en el Laboratorio de Tiempo Real del CIC - IPN.

Furui Sadaoki. Recent Advances in Speaker Recognition. Lectures Notes in Computer Science 1206 Springer. 1997.

Hayawaka Shoji. Speaker Identification Using Harmonic Structure of LP - Residual Spectrum. Lecture Notes in Computer Science 1206. Springer. 1997.

Kay Elemetrics, Corp. DSP Sona-Graph Mod. 5500-1. Workstation. Computer Speech Lab. 1998

Mansfred R. Schooeder. Relation between Cepstrum and Predictor Coefficients. Vol. ASSP – 29 No.2 april 1981

Neurofisiología Otooftalmológica BUENOS AIRES, ARGENTINA, 1998] http://vertigo-dizziness.com/html/quienes.html

Rabiner L. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, Vol. 77, No.2, feb 1989.

Rabiner L. & Juang B-H. Fundamentals of Speech Recognition. Prentice Hall, 1993

Reddy D.R.. An approach to Computer Speech Recognition by Direct Analysis of Speech Wave. Tech. Report No. C549. Computer Science Dept. Stanford Univ. Sept. 1966.

Reyes Carlos A.. On the Design of a Fuzzy Relational Neural Network for Automatic Speech Recognition Universidad del Estado de Florida, Colegio de Artes y Ciencias. Tesis Doctoral. 1994.

Savage Jesús. A Hybrid System with Symbolic AI and Statistical Method for Speech Recognition. Tesis Doctoral. Universidad de Washington. 1995.

Speech Wiever. IBM 1996.

Suárez Sergio. Procesamiento de voz, para aplicaciones médico foniátricas. Proyecto de investigación CIC-IPN, México D.F. Junio 2000.

Tinnitus International Journal" Instituto de Investigaciones Neurootológicas (4G-F) - Bad Kissingen-Alemania y Centro de Investigaciones del Tinnitus de la SUNY y de la Fundación Martha Entermann, New York, USA.

Wahlster W. Prof.. Verbmovil. Approximation Techniques for Spoken Dialog Understanding and Translation. IX International Symposium on Artificial Intelligence. Cancun, 13 nov. 1996.